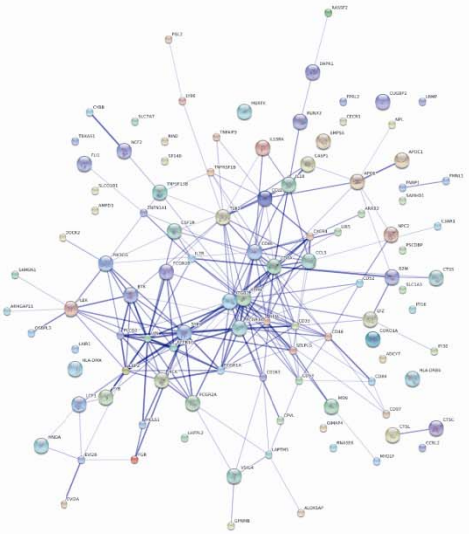
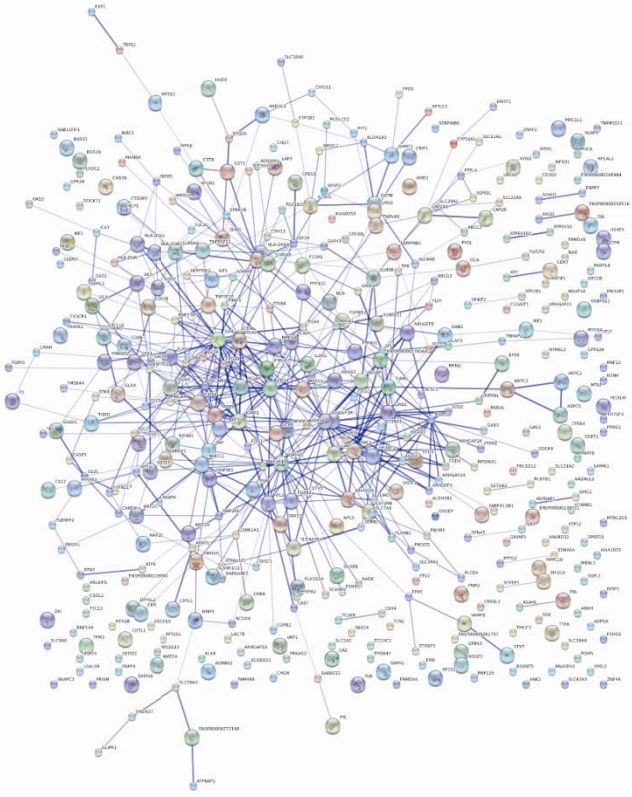
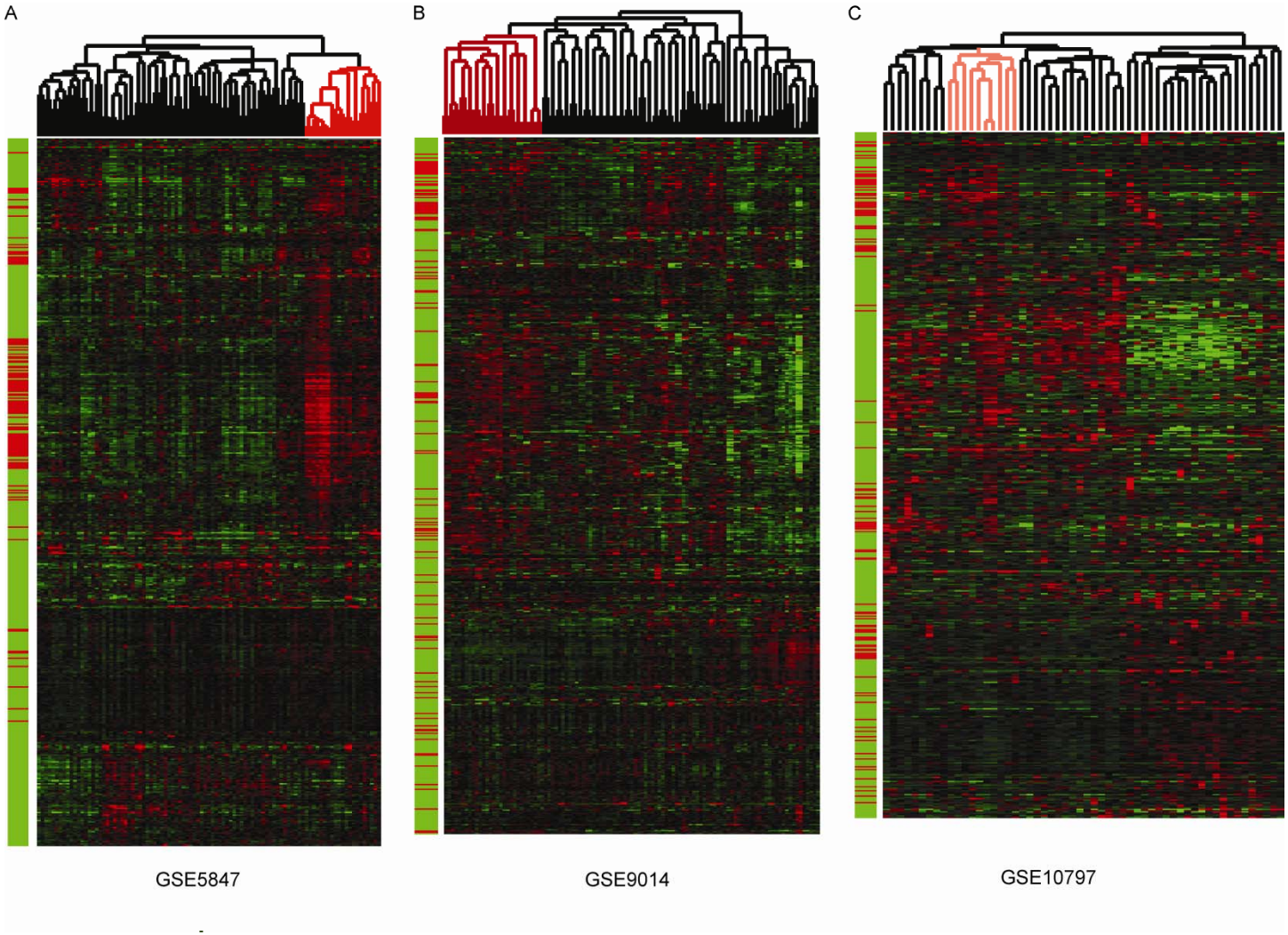
A CSF1 Core Protein-Protein Interaction Network



B CSF1 Non-Core Protein-Protein Interaction Network




A
B
C
GSE5847
GSE9014
GSE10797

Supplemental Methods:

Whole Tumor Breast Cancer Gene Expression Datasets:

We examined a total of five whole tumor breast cancer gene expression datasets with clinical outcome: NKI (1), Perreard (2), GSE1379 (3), GSE1456 (4), and GSE3494 (5). Gene expression and clinical data for GSE1379, GSE1456, and GSE3494 can be downloaded from the Gene Expression Omnibus (6). Gene expression and clinical data for the NKI dataset can be downloaded from the supplemental information from Chang *et al.* (http://microarray-pubs.stanford.edu/wound_NKI/) (7), and gene expression and clinical data for the Perreard dataset can be downloaded from the UNC Microarray Database (https://genome.unc.edu/cgi-bin/SMD/publication/viewPublication.pl?pub_no=48&1740).

NKI:

The NKI data set contains gene expression profiles from 295 cases of breast carcinoma, measured on a 25,000 spot oligonucleotide arrays as described in van de Vijver *et al*. (1). All patients included in this dataset were younger than 53 years old and had stage I or II disease.

Perreard Dataset:

The Perreard dataset contains a total of 126 samples of normal breast tissue, breast cancers, fibroadenomas, metastatic breast cancer, and breast cancer cell lines whose gene expression was measured on Agilent microarrays (Agilent Human A1, Agilent Human A2, and custom oligonucleotide) as described in Perreard *et al.* (2). The cases span an ethnically diverse cohort of patients from the University of Utah Health Sciences Center, University of North Carolina, Thomas Jefferson University, Maine Medical Center, and University of Chicago. Patients were treated in accordance with the standard of care dictated by their disease stage, ER status, and HER2 status. Patient outcome information was collected for up to 8.8 years with a

median of 1.8 years. Prior to downloading the data from the UNC data set, the following filtering criteria were used: only spots with a regression correlation greater than 0.6, channel 1 and 2 IN mean / BN median greater than 1.5, channel 1 and 2 lowess normalized net mean greater than or equal to 30 were used, and only cases of invasive breast carcinoma were analyzed. Arrays and genes with greater than 70% of data fulfilling these criteria were evaluated, and genes and arrays were centered by mean expression level. This resulted in 13,142 genes and 91 cases of invasive breast carcinoma. The provided clinical data for these cases included: age, ER status, lymph node status, grade, size, disease free survival, and overall survival.

GSE1379:

The GSE1379 data set contains gene expression data from 60 patients with hormone receptor-positive breast cancer diagnosed at the Massachusetts General Hospital and treated with standard breast surgery and radiation followed by five years of systemic adjuvant tamoxifen, as described in Ma *et al*. (3). Patient outcome information was collected for up to 14.1 years with a median of 7.3 years. The gene expression profiles were measured on a custom-designed Agilent 22,000-gene oligonucleotide microarray. The clinical information includes tumor type, size, grade, lymph node status, ER, PR, Her2, age, and disease free survival.

GSE1456:

The GSE1456 data set contains gene expression data from 159 patients operated on at the Karolinska Hospital and identified from the population-based Stockholm Gotland breast cancer registry, as described in Pawitan *et al*. (4). The outcome information was collected for up to 8.5 years with a median of 7.1 years. The gene expression profiles were measured on Affymetrix

U133 A arrays. The clinical data available for each patient includes: relapse free survival, breast cancer specific survival, histologic grade, and molecular subtype.

GSE3494:

The GSE3494 data set contains gene expression profiles of 251 patients who underwent resection of breast cancers in Uppsala County, Sweden, as described in Miller *et al*. (5). The gene expression profiles were measured on Affymetrix U133 A. Clinical data for each case includes: age, size, lymph nodes, breast cancer specific survival, p53 mutation status, and p53 DLDA gene signature status.

Breast Cancer Laser Capture Microdissection (LCM) Gene Expression Datasets:

We examined three breast cancer LCM datasets: GSE5847 (8), GSE10797 (9), and GSE9014 (10). Gene expression and clinical data for these three datasets can be downloaded from the Gene Expression Omnibus (6).

Antibodies and Scoring Technique Used for Immunohistochemistry:

For immunohistochemistry, the primary antibodies used were FCGR3a (CD16) (MCA1816, mouse monoclonal, AbD Serotec, CA), CTSL1 (MCA2374, mouse monoclonal, AbD Serotec, CA), FCGR2a (CD32) (AB45143, rabbit monoclonal, Abcam, UK), CD163 (NCL-CD163, mouse monoclonal, Novocastra, CA). The immunohistochemical reactions were visualized using mouse and rabbit versions of the EnVision + system (DAKO, Carpinteria, CA) using diaminobenzidine. CD163 staining was performed with the Ventana Benchmark Autostainer. *In situ* hybridization of TMA sections was performed based on a protocol published previously (11-13).

Stains were interpreted according to the following criteria: score 0 (negative) for < 10 positively stained cells, score 1 (weak) for ≥10 but <20 positively stained cells, score 2 (moderately) for ≥20 but <45 positively stained cells, and score 3 (strong) for ≥ 45 positively stained cells. For the quantification of CSF1 in situ hybridization, three scores were used: score 0 (negative) if no dots were seen in the cells, score 2 (weak) if there were few dots in a minority of cells, and score 3 (strong) if there were several dot in the majority of the cells. Cores in which no diagnostic material was present were omitted from further analysis. Following completion of the histopathological scoring, the numerical scores were processed using the Deconvoluter software as previously described (14), with each sample receiving the highest score for either of the multiple cores taken from each case.

References

1. van de Vijver MJ, He YD, van't Veer LJ, *et al.* A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347: 1999-2009.
2. Perreard L, Fan C, Quackenbush JF, *et al.* Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay. Breast Cancer Res 2006;8: R23.
3. Ma XJ, Wang Z, Ryan PD, *et al.* A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer Cell 2004;5: 607-16.
4. Pawitan Y, Bjohle J, Amler L, *et al.* Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. Breast Cancer Res 2005;7: R953-64.
5. Miller LD, Smeds J, George J, *et al.* An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. Proc Natl Acad Sci U S A 2005;102: 13550-5.
6. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30: 207-10.
7. Chang HY, Nuyten DS, Sneddon JB, *et al.* Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci U S A 2005;102: 3738-43.
8. Boersma BJ, Reimers M, Yi M, *et al.* A stromal gene signature associated with inflammatory breast cancer. Int J Cancer 2008;122: 1324-32.
9. Casey T, Bond J, Tighe S, *et al.* Molecular signatures suggest a major role for stromal cells in development of invasive breast cancer. Breast Cancer Res Treat 2008.
10. Finak G, Bertos N, Pepin F, *et al.* Stromal gene expression predicts clinical outcome in breast cancer. Nat Med 2008;14: 518-27.
11. St Croix B, Rago C, Velculescu V, *et al.* Genes expressed in human tumor endothelium. Science 2000;289: 1197-202.
12. Iacobuzio-Donahue CA, Ryu B, Hruban RH, Kern SE. Exploring the host desmoplastic response to pancreatic carcinoma: gene expression of stromal and neoplastic cells at the site of primary invasion. The American journal of pathology 2002;160: 91-9.
13. West RB, Corless CL, Chen X, *et al.* The novel marker, DOG1, is expressed ubiquitously in gastrointestinal stromal tumors irrespective of KIT or PDGFRA mutation status. The American journal of pathology 2004;165: 107-13.
14. Liu CL, Prapong W, Natkunam Y, *et al.* Software tools for high-throughput analysis and archiving of immunohistochemistry staining data obtained with tissue microarrays. Am J Pathol 2002;161: 1557-65.

Supplemental Table 1. CSF1 Response Core Gene Set and Non-Core Gene Set PPI Network Characteristics*

	CSF1 Response Core Network	CSF1 Response Non-Core Network	P value
Mean Clustering Coefficient	0.75	0.35	1.28E-05
Mean Degree	13.2	8.8	0.007
Mean Neighborhood Connectivity	8.2	5.9	0.002
Mean Closeness Centrality	0.39	0.31	0.002
Connected Components	3	13	N/A
Network Centralization	0.26	0.1	N/A

*The table compares the topological characteristics of the PPI networks created with 80 CSF1 response core proteins (CSF1 Response Core Network) and 285 CSF1 response non-core proteins (CSF1 Response Non-Core Network).

For a full explanation of the mathematical derivation and meaning of the network features, see http://med.bioinf.mpi-inf.mpg.de/netanalyzer/help/2.5.2/index.html#simple.

Supplemental Table 2. The CSF1 Response and Clinicopathologic Features in Gene Expression Analysis*

	CSF1 Response Signature Positive	CSF1 Response Signature Negative	P value
Estrogen receptor negative	43%	14%	1.77E-15
Progesterone receptor negative	41%	18%	6.83E-05
Histologic Grade			6.92E-10
Grade 1	12%	26%	
Grade 2	36%	46%	
Grade 3	53%	29%	
Mean Size (mm)	25	22	0.009
Mean age (years)	53	54	0.428
Lymph node metastasis	44%	42%	0.667
Mean Disease Free Survival, n=294 (years)	7.7	9.9	0.144
Mean Disease Specific Survival, n=395 (years)	10.5	10.7	0.691
Mean Overall Survival, n=529 (years)	11.6	14.2	0.099
Molecular Characteristics			
P53 mutation	46%	16%	2.40E-06
P53 mutation expression signature	61%	19%	3.40E-10
Basal molecular subtype	30%	11%	6.50E-06
ERBB2	22%	9%	5.00E-04
Normal-like	2%	13%	0.002
Luminal A	8%	11%	0.286
Luminal B	11%	7%	0.122

* Clinicopathological data pooled from 5 datasets (total n = 856).
The percentages refer to the proportion of the CSF1 response positive tumors or the CSF1 response negative tumors with a given clinicopathological feature.

Supplemental Table 3. The CSF1 Response and Clinicopathologic Features in TMA Analysis*

	CSF1 Response Signature Positive	CSF1 Response Signature Negative	P value
ER negative	45%	10%	6.40E-09
PR negative	48%	23%	0.0002
Histologic Grade			6.05E-12
Grade 1	6%	35%	
Grade 2	38%	50%	
Grade 3	56%	15%	
MeanKi67%	52%	26%	2.61E-07
EGFR expression	31%	6%	9.13E-08
Histologic Subtype			0.0003
Ductal	98%	79%	
Lobular	2%	22%	
Lymph Node Metastasis	43%	43%	0.999
Mean Tumor Size (mm)	24	24	0.904
Her 2 ratio >2.2	10%	10%	0.974

Total n with valid data = 206.

The percentages refer to the proportion of the CSF1 response positive tumors or the CSF1 response negative tumors with a given clinicopathological feature.